

The Netflix logo, consisting of the word "NETFLIX" in a bold, red, sans-serif typeface. The letters are slightly irregular and slanted, giving it a hand-drawn or blocky appearance. The logo is centered within a dark gray rectangular background.

# Project 1: Netflix Data

Group 11: Megan Palma, Elizabeth Bradshaw, Adrianna Johnson, Arman Bains, Amit Suresh

## Target Audience / Value Proposition

Netflix and other streaming services are the target audience for understanding trends in content genre, ratings, movie run time and seasons to find themes that connect with their audience to ensure their content library is relevant.

## Motivations

Our motivation was to gain an understanding of the trends in content published on Netflix from the past 20 years and to help streaming services better understand the market. At the end we will make a few recommendations.

## Tools Used

Pandas, Matplotlib, NumPy, and Script.Stats



# Introduce Data / Sources

- Data Source: Bansal, Shivam. “Netflix Movies and TV Shows.” *Kaggle*, [www.kaggle.com/datasets/shivamb/netflix-shows](https://www.kaggle.com/datasets/shivamb/netflix-shows). Accessed 13 Mar. 2024.
- **Data Represents:**
  - Original Data Consisted of 12 Columns of ‘Show ID’, ‘Type’, ‘Title’, ‘Director’, ‘Cast’, ‘Country’, ‘Date Added’, ‘Release Year’, ‘Rating’, ‘Listed In’, and ‘Description’
  - Each data set represents a Movie and TV Show on Netflix
  - While reviewing this original data set we decided to focus on Type, Genre, Rating, and Release Year for the Top 10 Countries
    - Decided to separate the data between Movies and TV Shows

[illegible]

# Data Processing Pipeline



## Cleaning data included:

- Removing rows with null values in columns 'country' and 'rating'
- Limiting dataframe to relevant columns IE 'type', 'country', 'rating', etc
- Splitting 'listed\_in' columns (genres) and 'date\_added' (date added by year)
- Reducing rows to include only the Top 10 countries by the number of productions in both Movies and TV Shows

0	Movie	Dick Johnson Is Dead	United States	September 25, 2021	2020	PG-13	90 min	Documentaries	September 25	2021	...	0	0	0	0	0	0	0	0	0
4	TV Show	Kota Factory	India	September 24, 2021	2021	TV-MA	2 Seasons	International TV Shows, Romantic TV Shows, TV ...	September 24	2021	...	1	0	0	0	0	0	0	0	0
8	TV Show	The Great British Baking Show	United Kingdom	September 24, 2021	2021	TV-14	9 Seasons	British TV Shows, Reality TV	September 24	2021	...	0	0	0	0	0	0	0	0	0
9	Movie	The Starling	United States	September 24, 2021	2021	PG-13	104 min	Comedies, Dramas	September 24	2021	...	0	0	0	0	0	0	0	0	0
15	TV Show	Dear White People	United States	September 22, 2021	2021	TV-MA	4 Seasons	TV Comedies, TV Dramas	September 22	2021	...	1	1	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

# Data Processing Pipeline

- Splitting 'listed\_in' columns (genres)

```
#breaking down the genres
categories_df=Cleaned_Data['listed_in'].str.get_dummies(',')
categories=list(categories_df.columns)
display(len(categories))
Cleaned_Data[categories]=categories_df
Cleaned_Data['num_categories']=categories_df.sum(axis=1)
Cleaned_Data
```

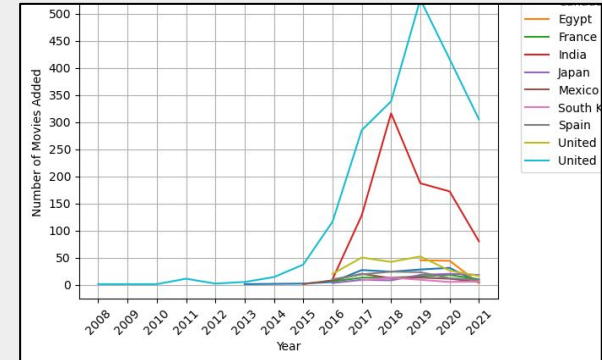
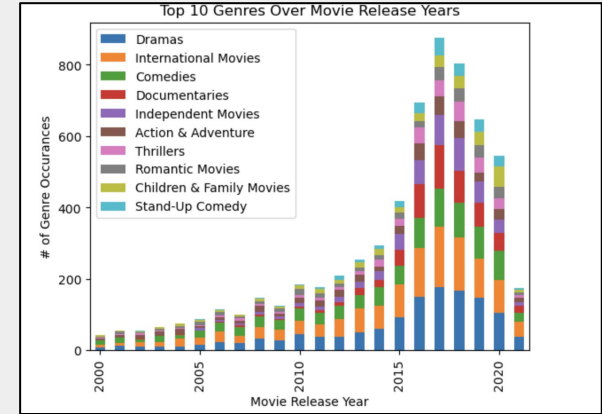
42

- Reducing rows to include only the Top 10 countries by the number of productions in both Movies and TV Shows.

```
#Movies and TV Shows produced in Countries by size. Top 10.
newt = Cleaned_Data.groupby('country').size().sort_values(ascending=False).head(10)
newt = newt.index.tolist()
Top_10_df = Cleaned_Data[Cleaned_Data['country'].isin(newt)]
Top_10_df
```

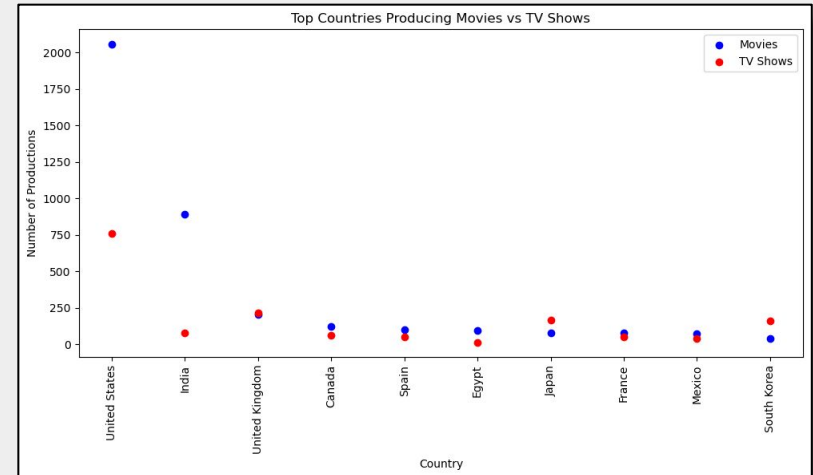
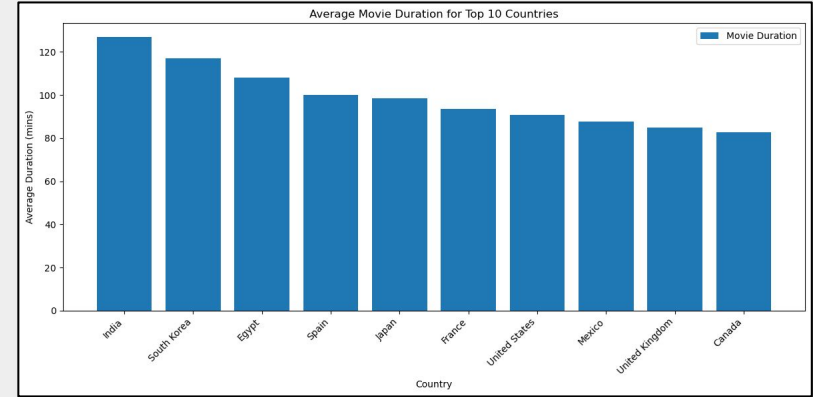
# Popular Genres & Content Added

- **Genres Analysis:**
  - 2 most popular genres across (both movies and TV Shows)
    - Drama and International Shows / Movies
    - This is not surprising as the Drama Genre is considered the top genre for streaming services based on the [Source](#)
  - Overlapping popular genres (both movies and TV shows)
    - Drama, International, Comedy, Documentary/Docuseries, Romance, and Kids/Family
- **Movies Added Per Year:**
  - The United States added most, followed by India
  - All amounts of movies added from 2019 to 2021 hit a decline due to the Covid-19 pandemic production shutdown.



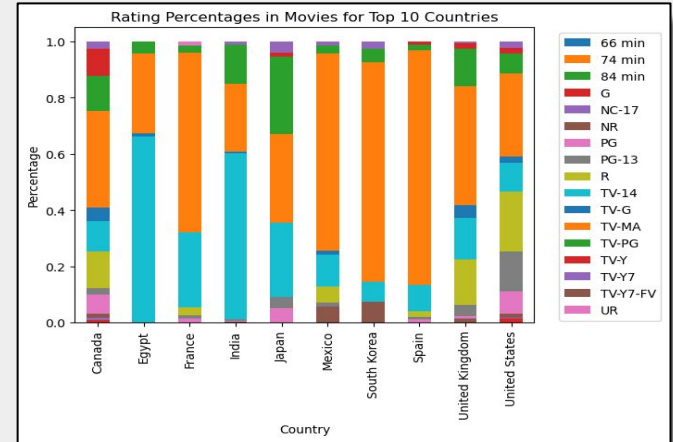
# Production & Duration

- **Production Analysis:**
  - The United States and India produced the most combined movies and tv shows with 2800 titles and 972 titles respectively
  - The top five countries include the UK, Canada, and Japan alongside India and the US
  - Outside of the UK, Japan, and South Korea, films are produced at a higher rate
- **Show Seasons / Movie Duration:**
  - Indian movies average the longest duration among top ten countries, while Canada is the shortest
  - Contrastly, Canada averages the most seasons among top producing countries, with Egypt being the least



# Content Ratings: Movies

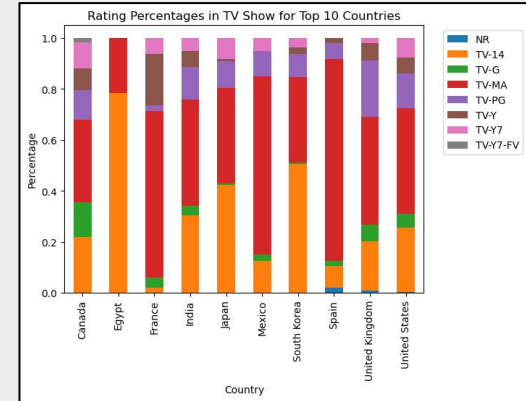
- Rating Percentage for Movies
- For top 10 countries the US and India carried the largest proportion of movie ratings at 55.19% and 23.94% respectively
  - Large portion of India's rating came from TV-14
  - TV-MA and R accounted for most ratings in US
- TV-MA was very predominant for most of the countries
- R movies were very low or virtually non-existent
- Except of course in the US where they accounted for 11.80% of the entire data





# Content Ratings: TV Shows

- Rating Percentage for TV Shows
- The United Kingdom, Japan and South Korea had strong showings for TV ratings
  - The ratings were 13.41%, 10.57% and 9.95% respectively for those nations.
  - The US was still on top at 47.86%, but the tv ratings compared to other countries was more competitive unlike for movies
- TV-MA and TV-14 were the high scorers across the top ten nations
- It seems a few countries like Japan and South Korea invest more resources to their TV industry
  - Both countries drew large success from the TV-14 rating



country	Canada	Egypt	France	India	Japan	Mexico	South Korea	Spain
rating								
NR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
TV-14	0.82	0.69	0.06	1.51	4.47	0.31	5.04	0.25
TV-G	0.50	0.00	0.13	0.19	0.06	0.06	0.06	0.06
TV-MA	1.20	0.19	2.02	2.08	3.97	1.76	3.34	2.39
TV-PG	0.44	0.00	0.06	0.63	1.13	0.25	0.88	0.19
TV-Y	0.31	0.00	0.63	0.31	0.06	0.00	0.25	0.06
TV-Y7	0.38	0.00	0.19	0.25	0.88	0.13	0.38	0.00
TV-Y7-FV	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00
country	United Kingdom		United States					
rating								
NR	0.13		0.13					
TV-14	2.58		12.15					
TV-G	0.88		2.52					
TV-MA	5.67		19.84					
TV-PG	2.96		6.61					
TV-Y	0.94		2.96					
TV-Y7	0.25		3.65					
TV-Y7-FV	0.00		0.00					

# Recommendations

- Netflix can utilize this data by take 2 approaches:
  - Continue the trend they are currently on by adding more of it's top 10 genres, run times / seasons, and further leveraging the top 10 countries producing countries
  - Further analyze trends across all streaming services and cater to any gaps



# Challenges and Next Steps

## Challenges:

- Original data wasn't completely accurate due to some data being in the wrong column
  - IE Ratings column had some duration data in it
- Duration Column was hard to separate between TV Shows seasons and Movie Duration
- Due to the large amount of data we had to be very selective on what we included
  - IE we slimmed down to the top 10 Countries and for Genres we further slimmed down to top 10 Genres

## Next Steps:

- Look at data for other streaming services
- Further analysis: the data on any trending content creator by leveraging the director and actor columns



# NETFLIX

Thank you!